THE DAY
FIND YOUR VOICE

Social media giants to label AI-made content

Reading Level 4

Is it enough? Experts warn that AI-generated deepfakes could affect the outcome of crucial upcoming elections. But measures are finally being taken to combat them.

Fakery: AI-generated images of Donald Trump have already been spread widely on social media.

Jolene looks at the video in amazement. There is Donald Trump handing out food to hungry children; giving a bed in his own house to a homeless man; addressing a rally in support of women's rights; receiving a medal for courage in Vietnam; punching President Putin on the nose! Maybe she should vote for him in the presidential election.

But then she notices a small label in the corner of her screen: "Made with **AI** ."

Meta announced last week that, starting in May, labels like this would appear on Instagram, Facebook and Threads. They will be applied to AI-generated videos, pictures and audio clips.

The power of AI to make mischief is already clear. There have been fake audio clips of **Keir Starmer** and **Sadiq Khan**, and fake phone calls from President Biden. There has even been a whole fake interview with a Ukrainian official appearing to claim credit for the recent terrorist attack in Moscow.

Many people are concerned about the influence such posts could have on upcoming elections. The question is whether effective countermeasures can be implemented in time. America's presidential election will take place in November, and Britain's general election by the end of January.

In February, tech companies including TikTok, Microsoft and X signed an agreement promising to combat political fakes. **New Hampshire**'s attorney general announced legal action against the company which faked the Biden phone calls.

That same month the US **House of Representatives** set up a taskforce to see how AI might be regulated. In March, leading researchers signed an open letter calling for AI companies to be made legally responsible for harmful fakes using their technology.

Both Meta and **OpenAI** have started putting **watermarks** on AI-generated images. A new charity, TrueMedia.org, has released free tools to help anyone detect fakes.

But the charity's head, Dr Oren Etzioni, warns that no tools are entirely effective. When it comes to elections, "I'm terrified. There is a very good chance we are going to see a **tsunami** of misinformation."[1]

To deal with it, there needs to be co-operation between governments, AI companies and tech giants — and he does not believe the chances of that happening before November are high.

Is it enough?

Fakery snakery

Yes: People are on the lookout for deepfakes anyway, and this will make them far easier to spot. Tech giants like Microsoft have so much expertise at their disposal that little will get past them.

No: TThe technology used to create deepfakes is amazing and is going to get more and more sophisticated as time goes on. It will always be one step ahead of those who are trying to combat it.

Or... It is unclear whether deepfakes can swing an election or not. China's attempts to undermine Taiwan's recent elections failed, but Russia's efforts in Slovakia seem to have been successful.

1. *Quoted in The New York Times.*

Key words

AI:
A computer programme that has been designed to think.

Meta: The new name of the company which owns Facebook and Instagram.

Keir Starmer: The leader of the UK Labour Party since 2020.

Sadiq Khan:
The current Mayor of London.

New Hampshire:
A state in the north-east US with a population of nearly 1.4 million.

House of Representatives:
The lower chamber of the United States congress. There are 435 representatives, with a certain number allocated to each state based on the state's population.

OpenAI: An American artificial intelligence company. It says its mission is to "benefit all of humanity".

Watermarks: A logo, text or pattern that is deliberately put over an image to make it more difficult to use without permission.

Tsunami: A Japanese word describing a succession of waves caused when an earthquake or volcano displaces a large body of water.

Six steps to discovery

1 Connect
How do you feel about this story?

Do you worry about deepfakes? Would you mind one being made of you?

2 Wonder
What questions do you have?

For example: How were deepfakes invented? What other elections might be affected?

3 Investigate
What are the facts?

Follow the link to the Guardian podcast to discover how deepfake videos are made.

4 Construct
What is your point of view?

Your MP asks you whether a forthcoming election can be conducted fairly. Think about what you would say.

5 Express
What do others believe?

Should their be prison sentences for people who create harmful deepfakes? Hold a class debate.

6 Reflect
What might happen next?

Imagine you discover that a deepfake has been made of your best friend. Write a story about it.

Some people say

"The great majority of mankind are satisfied with appearances... and are often more influenced by the things that seem than by those that are."

Niccolò Machiavelli (1469 – 1527), Italian diplomat

"Those who think their intellect will keep them from deception are already deceived."

Bill Johnson (1951 –), American clergyman

What do you think?

Dive in deeper

A convincing deepfake in support of Donald Trump. TrueMedia.org (0:17)

An eye-opening video about how TrueMedia.org works to detect fakes. TrueMedia.org (2:03)

A news report on Meta's announcement. Reuters (400 words)

An authoritative article about China's use of deepfakes. The Guardian (600 words)